

Федеральное государственное бюджетное образовательное учреждение высшего образования Московский государственный университет имени М.В. Ломоносова  
ФИЛИАЛ МГУ В Г. ДУБНЕ



УТВЕРЖДАЮ  
И.о. директора  
филиала МГУ в г.Дубне  
/ Э.Э. Боос /  
\_\_\_\_\_ 09 \_\_\_\_\_ 2022 г.

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

**Наименование дисциплины:**

Современные методы анализа данных

---

**Уровень высшего образования:**

Магистратура

---

**Направление подготовки:**

03.04.02 Физика

---

**Направленность (профиль) ОПОП:**

Физика элементарных частиц

---

**Форма обучения:** Очная форма обучения

---

Дубна 2022 г.

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки программы магистратуры 03.04.02 «Физика».

Год (годы) приема на обучение \_\_\_\_\_

**Авторы–составители:**

1. Кандидат физ.-мат. наук Леонтьев Владимир Викторович, доцент кафедры физического факультета МГУ

Руководитель магистерской программы:

Доктор физ.-мат. наук, профессор академик РАН В.А. Матвеев, заведующий кафедрой физического факультета МГУ

## **Аннотация к рабочей программе дисциплины «Современные методы анализа данных»**

Целью данного курса является подготовка студентов к решению ряда типичных задач, которые ставятся начинающим научным сотрудникам в современной физике атомного ядра и частиц, а именно: применение программных средств для анализа и последующего представления научных данных, либо полученных в эксперименте, либо произведенных на основе теоретических моделей. В курсе содержатся базовые знания о принципах работы таких программ, а также о необходимом аппаратно-программном инструментарии для их работы. В рамках курса студенты получают базовые навыки для работы с рядом современных программных пакетов анализа данных и представления полученных результатов, ознакомятся с их характеристиками и областями применения.

### **Разделы рабочей программы**

1. Место дисциплины в структуре основной профессиональной образовательной программы высшего образования (ОПОП ВО).
2. Входные требования для освоения дисциплины (модуля), предварительные условия (при наличии)
3. Результаты обучения по дисциплине (модулю), соотнесенные с формируемыми компетенциями
4. Формат обучения.
5. Язык обучения.
6. Содержание дисциплины.
7. Объем дисциплины
8. Структурированное по темам (разделам) содержание дисциплины (модуля) с указанием отведенного на них количества академических часов и виды учебных занятий
9. Текущий контроль и промежуточная аттестация.
10. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю).
11. Шкала оценивания.
12. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости.
13. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.
14. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

## 1. Место дисциплины в структуре ОПОП ВО

Дисциплина «Современные методы анализа данных» реализуется на 1-ом курсе во 2-ом семестре магистратуры и входит в состав вариативной части.

## 2. Входные требования для освоения дисциплины (модуля), предварительные условия

Базовые знания в области общей и теоретической физики в объеме классических университетских курсов

## 3. Результаты обучения по дисциплине (модулю), соотнесенные с формируемыми компетенциями

В результате освоения дисциплины у обучающихся должны быть сформированы:

Код и наименование компетенции	Код и наименование индикатора (индикаторов) достижения компетенции	Результаты обучения
<b>МПК-2</b> Способен ставить, формализовать и решать задачи в области физики элементарных частиц	<b>ИМПК-2.3</b> Знает основные и перспективные направления физики элементарных частиц	<b>З-1</b> Знать: особенности Unix-подобных ОС, способы взаимодействия с ними, в том числе в случае с удаленным доступом <b>У-1</b> Уметь: осуществлять моделирование и анализ событий физики частиц, используя существующие программные пакеты и методы <b>В-1</b> Владеть: навыками программирования C++, Python и иных языков и дополнительных библиотек, достаточными для использования методов в ROOT или иных пакетах анализа

2. **Форма обучения:** очная.

3. **Язык обучения:** русский.

4. **Содержание дисциплины**

*Тема 1. Unix-подобные ОС.*

Необходимый инструментарий для работы с программами анализа: Unix-подобные ОС. Режим удаленного терминала, файловая система, разграничение доступа к ресурсам, работа с файлами и процессами, настройки shell и т.п.

*Тема 2. Средства коммуникации.*

Необходимый инструментарий для работы с программами анализа: средства коммуникации. Почтовые клиенты, средства WWW, протоколы копирования файлов. Доступ к Unix/Linux-серверу с клиентской Windows-машины.

*Тема 3. Введение в программирование.*

Необходимый инструментарий для работы с программами анализа: введение в программирование на C++ в Unix-среде. Сборка программ, утилита make, управление программным потоком ввода/вывода.

*Тема 4. Пакет ROOT, часть 1.*

Анализ и представление экспериментальных данных (настройка переменных окружения, гистограммы, фитирование, графические возможности ROOT, экспорт/импорт.

*Тема 5. Пакет ROOT, часть 2.*

Графы, коллекции объектов TTree, итераторы.

*Тема 6. Моделирование и анализ физических событий высоких энергий.*

Жесткие процессы физики элементарных частиц. Пакет Pythia для моделирования событий физики высоких энергий.

*Тема 7. Моделирование и анализ физических событий при промежуточных энергиях.*

Адронная физика, многоступенчатые распады частиц, неизотропные распределения. Пакет PLUTO.

*Тема 8. Анализ отклика детектирующих систем.*

Анализ отклика детектирующих систем в физике высоких энергий. Описание геометрии детектора, описание взаимодействия частиц и излучения с веществом детектора, моделирование отклика детектора, сравнение с данными эксперимента, программы визуализации.

**7. Объем дисциплины**

НАЗВАНИЕ ДИСЦИПЛИНЫ	Трудоемкость в зачетных единицах	объем учебной нагрузки в ак. часах					Самостоятельная работа студентов
		Общая трудоемкость	в том числе ауд.занятий			Учебно-практические занятия	
			Общая аудиторная нагрузка	Лекций	Семинаров		
Современные методы анализа данных	2	72	34	17	17	38	

**8. Структурированное по темам (разделам) содержание дисциплины (модуля) с указанием отведенного на них количества академических часов и виды учебных занятий**

Изучение курса «Современные методы анализа данных» включает в себя лекции, на которых рассматривается теоретическое содержание курса; семинарские занятия, предусматривающие углубленное изучение и обсуждение вопросов, обозначенных в темах дисциплины; самостоятельную работу, заключающуюся в подготовке к лекционным и семинарским занятиям. Темы, рассматриваемые на лекциях и изучаемые самостоятельно, закрепляются на семинарских занятиях, по вопросам, вызывающим затруднения, проводятся консультации.

№ тем ы		Виды учебной нагрузки и их трудоемкость, часы					Са мо ст оя те ль на я ра бо та	Форма текущего контроля успеваемости и промежуточной аттестации
		Вс его ча со в	Л е к ц и и	С е м и н а р ы	Учебно-практические занятия (лабораторные или практические занятия)			
1	Unix-подобные ОС	8	2	2		4		
2	Средства коммуникации	8	2	2		4		
3	Введение в программирование	8	2	2		4		
4	Пакет ROOT, часть 1	8	2	2		4		
5	Пакет ROOT, часть 2	8	2	2		4		
6	Моделирование и анализ физических событий высоких энергий	8	2	2		4		
7	Моделирование и анализ физических событий при промежуточных энергиях	8	2	2		4		
8	Анализ отклика детектирующих систем	12	3	3		6		
	Промежуточная аттестация	4				4	зачет	
<b>ИТОГО:</b>		72	17	17		38		

ДЗ- домашнее задание, Оп- опрос, КР- контрольная работа, Т- тестирование, Реф - реферат

## 9. Текущий контроль и промежуточная аттестация.

Текущий контроль по дисциплине «Современные методы анализа данных» осуществляется на лекциях и семинарских занятиях и заключается в оценке активности и

качества участия в опросах и собеседованиях по проблемам, изучаемых в рамках тем лекционных занятий, аргументированности позиции; оценивается ширина используемых теоретических знаний.

Промежуточная аттестация по дисциплине «Современные методы анализа данных» проводится во втором семестре в форме зачета в виде письменной работы.

#### 10. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю).

Перечень оценочных средств, применяемых на каждом этапе проведения текущего контроля и промежуточной аттестации по дисциплине, представлены в таблице

Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде
<b>Оценочные средства текущего контроля</b>		
Тематический опрос (в форме ответов на вопросы)	Средство контроля, организованное как специальная беседа по тематике предыдущей лекции и рассчитанное на выяснение объема и качества знаний, усвоенных обучающимися по определенному разделу, теме, проблеме.	Перечень тем, изучаемых в рамках дисциплины
Собеседование (в форме беседы, дискуссии по теме)	Средство контроля, организованное как свободная беседа, дискуссия по тематике изучаемой дисциплины, рассчитанное на выяснение объема знаний обучающегося по всем изученным разделам, темам; свободного использования терминологии для аргументированного выражения собственной позиции.	Перечень тем, изучаемых в рамках дисциплины
Тестирование	Средство контроля, позволяющее получить оценку уровня фактических знаний аспиранта по изученной теме.	Образцы тестов
<b>Оценочные средства промежуточной аттестации</b>		
Письменная работа	Средство, позволяющее оценить сформированность систематических представлений о методах научно-исследовательской деятельности.	Перечень вопросов
Собеседование	Средство, позволяющее получить экспертную оценку знаний, умений и навыков по для оценивания и анализа различных фактов и явлений в своей профессиональной области.	Требования к порядку проведения собеседования

#### 11. Шкала оценивания.

Планируемые результаты обучения	Критерии оценивания результатов обучения			
	2/ не зачтено	3/ зачтено	4/ зачтено	5/ зачтено
ЗНАТЬ:	Отсутствие знаний	В целом	В	Успешные и

особенности Unix-подобных ОС, способы взаимодействия с ними, в том числе в случае с удаленным доступом ИМПК-2.3	особенностей Unix-подобных ОС, способов взаимодействия с ними, в том числе в случае с удаленным доступом	успешные, но не систематические знания особенностей Unix-подобных ОС, способов взаимодействия с ними, в том числе в случае с удаленным доступом	целом успешное, но содержащее отдельные пробелы знания особенностей Unix-подобных ОС, способов взаимодействия с ними, в том числе в случае с удаленным доступом	систематические знания особенностей Unix-подобных ОС, способов взаимодействия с ними, в том числе в случае с удаленным доступом
УМЕТЬ: осуществлять моделирование и анализ событий физики частиц, используя существующие программные пакеты и методы ИМПК-2.3	Отсутствие умения осуществлять моделирование и анализ событий физики частиц, используя существующие программные пакеты и методы	В целом успешное, но не систематическое умение осуществлять моделирование и анализ событий физики частиц, используя существующие программные пакеты и методы	В целом успешное, но содержащее отдельные пробелы умение осуществлять моделирование и анализ событий физики частиц, используя существующие программные пакеты и методы	Успешное и систематическое умение осуществлять моделирование и анализ событий физики частиц, используя существующие программные пакеты и методы
ВЛАДЕТЬ: навыками программирования C++, Python и иных языков и дополнительных библиотек, достаточными для использования методов в ROOT или иных пакетах анализа ИМПК-2.3	Отсутствие/фрагментарное владение навыками программирования C++, Python и иных языков и дополнительных библиотек, достаточными для использования методов в ROOT или иных пакетах анализа	В целом успешное, но не систематическое владение навыками программирования C++, Python и иных языков и дополнительных библиотек, достаточными для использования методов в ROOT или иных пакетах анализа	В целом успешное, но содержащее отдельные пробелы владение навыками программирования C++, Python и иных языков и дополнительных библиотек, достаточными для использования методов в ROOT или иных пакетах анализа	Успешное и систематическое владение навыками программирования C++, Python и иных языков и дополнительных библиотек, достаточными для использования методов в ROOT или иных пакетах анализа

## 12. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости.

### *Материалы текущего контроля успеваемости обучающихся*

#### **Пример:**

#### Вопросы по теории:

1. Описываются ли процессы адронизации теорией малых возмущений?
2. Философия Монте-Карло метода

## 13. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.

Зачет проводится в форме письменной работы.



## *Материалы промежуточной аттестации обучающихся*

**Пример практического задания.** Анализ данных полупеттонового распада пары топ-антитоп кварков (калибровка на процессе Дрелла-яна)

1. Зарегистрируйтесь на компьютере, который имеет:
  - 1.1. Установленный пакетом ROOT 6 (на машине lx.msu.dubna.ru нужно сделать соответствующую настройку переменных shell);
  - 1.2. Установленный компилятор C++ версии 4.9+ (на Unix-подобных операционных системах можно проверить версию компилятора командой `gcc --version`, на машине lx.msu.dubna.ru нужно запустить пакет разработчика командой `scl enable devtoolset-3 bash`);
2. Скачайте и распакуйте архив с рассматриваемым примером `HEPTutorial_16.tgz` [1], зайдите в получившуюся директорию;
3. Ознакомьтесь с документацией к данному заданию «CMS data analysis tutorial»[1];
4. Сделайте тестовую компиляцию исполняемого файла анализа `example.x` и его запуск;
5. В данный момент программа создает гистограмму для отображения количества восстановленных, изолированных от адронных струй мюонов, но заполняет ее единицами (с весами для всех фоновых процессов). Файлы вывода для экспериментальных данных `results.pdf`, для Монте-Карло данных `results_MC.pdf`. Заполните эту гистограмму настоящими значениями.
6. В данный момент программа создает гистограмму для отображения инвариантной массы пары мюонов, но заполняет ее единицами. Заполните эту гистограмму настоящими значениями, при этом используйте только пару изолированных, противоположно заряженных мюонов. Для расчета инвариантной массы можно посмотреть методы класса `TLorentzVector`, либо пересчитывая из величин энергии и проекции импульса каждого мюона, либо из магнитуды суммы их 4-векторов.
7. Сравните результаты анализа экспериментальных данных и Монте-Карло моделирования
8. Обновленные файлы `results.pdf`, `results_MC.pdf`, а также код анализа `MyAnalysis.C` пришлите в качестве ответа на задание, веб-форму задания на сайте[1].

### **Контрольные вопросы к практическому заданию:**

1. Как часто на гистограмме для отображения количества восстановленных, изолированных от адронных струй мюонов проявляются события с наличием более одного изолированного мюона?
2. Откуда появляются дополнительные мюоны на данной гистограмме?

### Вопросы к зачету:

## **14. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы**

### Основная литература

1. «ROOT CSC-2010 Handout», Lectures at the CERN School of Computing (CSC-2010) [http://axel.web.cern.ch/axel/CSC/ROOT\\_CSC10-Handout.pdf](http://axel.web.cern.ch/axel/CSC/ROOT_CSC10-Handout.pdf)
2. В.В. Галактионов и др. «Руководство для пользователей Linux-кластера ОИЯИ», ОИЯИ, P11-2004. <http://www.info.jinr.ru/unixinfo/usersguide/>

### Дополнительная литература

1. I. Antcheva «ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization» Computer Physics Communications Volume 180, Issue 12, December 2009, Pages 2499-2512

2. J. Allison et al., «Geant4 Developments and Applications», IEEE Transactions on Nuclear Science 53 No. 1 (2006) 270-278.

При реализации дисциплины может быть использовано следующее программное обеспечение:

1. Операционная система Astra Linux ( <https://astralinux.ru/> ) или аналог, с офисным пакетом, с пакетами разработчика.

Перечень профессиональных баз данных и информационных справочных систем

1. База данных РИНЦ (российский индекс научного цитирования) <http://www.elibrary.ru>
2. <https://msu-dubna.ru/her> (Курсы/ Магистратура/1 курс/ Раздел «Информационные методы»). Область на официальном сайте кафедры физики элементарных частиц (г. Дубна), где расположены необходимые ресурсы для данного курса.
3. <http://home.thep.lu.se/~torbjorn/Pythia.html> Публичный сайт разработчика PYTHIA
4. <https://www-hades.gsi.de/pluto> Публичный сайт разработчика PLUTO
5. <http://opendata.cern.ch> Сайт публичных экспериментальных данных CERN

#### **Материально-техническое обеспечение**

В соответствии с требованиями п. 5.3. образовательного стандарта МГУ по направлению подготовки «Физика», курс может быть прочитан в поточной аудитории при наличии: работающих электрических розеток, компьютеров для каждого студента, проектора, экрана, учебной доски.